> NIH Chest X-ray14: Pathology Detection

1. Project Objective

The goal of this work is to train a convolutional neural network (CNN) model to perform multi-label classification of chest X-ray images, identifying diseases such as Pneumonia and Emphysema, among others, using the NIH Chest X-ray14 dataset.

2. Dataset Used

Source: NIH Chest X-ray14
Total images available (after processing): ≈112,000
Format: PNG
Labels: Multiclass (each image may contain one or more diseases)
Included labels: The main extracted classes were
['Atelectasis', 'Cardiomegaly', 'Effusion', 'Emphysema', 'Infiltration', 'Mass', 'Nodule', 'Pneumonia', 'Pneumothorax'].

3. Data Preprocessing

The .csv files containing labels were loaded and linked to their corresponding images located in multiple subdirectories.
Multiclass text labels were transformed into a binarized format using MultiLabelBinarizer().
Images without corresponding entries on disk were discarded.
A stratified split was applied:

- Training: 72%
- Validation: 8%
- Test: 20%

4. Image Preprocessing

Input size: (224, 224, 3)
Normalization: Rescaled to range [0,1]
Augmentation: (training only)
Additional transformations included rotation, zoom, horizontal and vertical shifts, and horizontal flips.

5. Model Architecture

Sequential CNN model composed of:

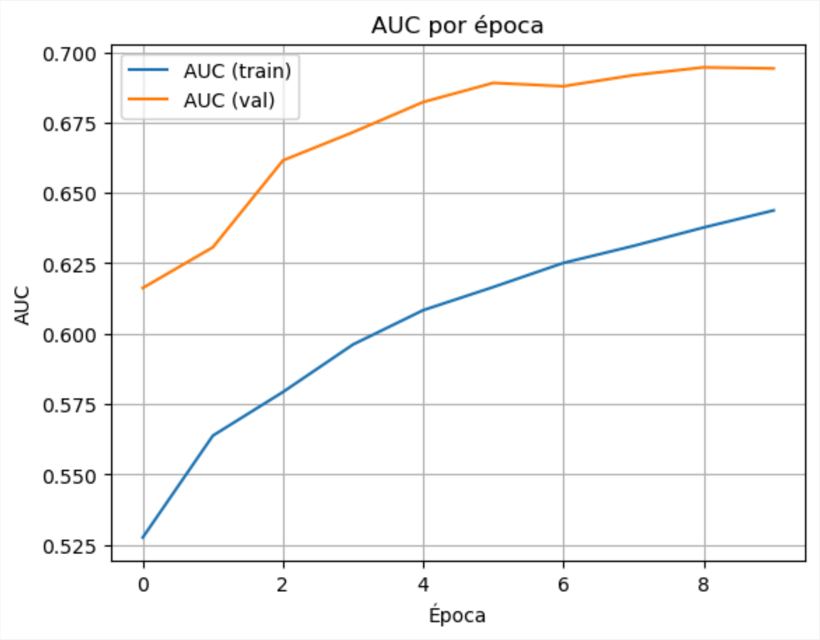
Layer	**Details**
Conv2D (32)	Kernel 3x3, ReLU activation
MaxPooling2D	2x2 pool size
Conv2D (64)	Kernel 3x3, ReLU activation
MaxPooling2D	2x2 pool size
Conv2D (128)	Kernel 3x3, ReLU activation
MaxPooling2D	2x2 pool size
Flatten	Volume flattening
Dense (128)	ReLU activation + 0.5 dropout
Dense (n_clases)	Sigmoid activation (multilabel output)

6. Compilation and Training

Optimizer: Adam (lr = 0.0001)
Loss function: binary_crossentropy (suitable for multi-label tasks)
Main metric: AUC (Area Under the ROC Curve)
Epochs: 10
Batch size: 64
Callbacks: ModelCheckpoint (saves the best model based on val_loss)

7. Training Results

After setting the input resolution to 224x224, a substantial improvement in the model's learning capacity was observed. The AUC curve shows a stable progression, with validation AUC values reaching around 0.70, indicating that the model successfully captures useful diagnostic patterns from the images.



8. Visual Evaluation

A subset of test images was visualized with their true and predicted classes.
A threshold of 0.5 was applied to determine class presence.
Each image shows:

- True: Ground-truth classes based on annotations
- Pred: Classes predicted by the model

This allows verification of whether the model correctly identifies multiple pathologies per image.

9. Conclusion and Observations

The model demonstrates the ability to detect multiple diseases in chest X-ray images using a supervised multi-label learning approach. The initial image size (32x32) limited the model's ability to capture subtle medical details. Using a 224x224 input size had a clearly positive impact on performance.
The gap between training and validation decreased, suggesting that the model is less constrained by input resolution. The model also shows no strong signs of overfitting, meaning it could still benefit from additional epochs or a deeper architecture.
The AUC metric is appropriate for evaluating multi-label models, particularly when classes are imbalanced. Finally, the validation vs. training curves show no strong signs of overfitting, which is a positive outcome.